LEVEL C

001922-2-T

social science research institute

RESEARCH REPORT

VALIDATION AND ERROR IN MULTIPLICATIVE
UTILITY FUNCTIONS

F. HUTTON BARRON

DECEMBER, 1978
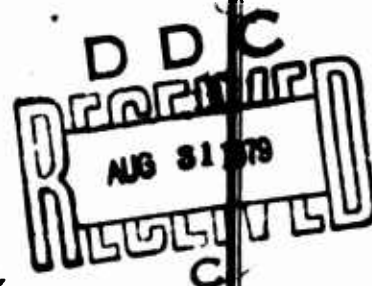
SSRI RESEARCH REPORT 78-2

79 08 31 058

Social Science Research Institute
University of Southern California
Los Angeles, California 90007
(213) 741-6955

INSTITUTE GOALS:

The goals of the Social Science Research Institute are threefold:

- To provide an environment in which scientists may pursue their own interests in some blend of basic and methodological research in the investigation of major social problems.

- To provide an environment in which graduate students may receive training in research theory, design and methodology through active participation with senior researchers in ongoing research projects.

- To disseminate information to relevant public and social agencies in order to provide decision makers with the tools and ideas necessary to the formulation of public social policy.

HISTORY:

The Social Science Research Institute, University of Southern California, was established in 1972, with a staff of six. In fiscal year 1978-79, it had a staff of over 90 full- and part-time researchers and support personnel. SSRI draws upon most University academic Departments and Schools to make up its research staff, e.g. Industrial and Systems Engineering, the Law School, Psychology, Public Administration, Safety and Systems Management, and others. Senior researchers have joint appointments and most actively combine research with teaching.

FUNDING:

SSRI Reports directly to the Executive Vice President of USC. It is provided with modest annual basic support for administration, operations, and program development. The major sources of funding support are federal, state, and local funding agencies and private foundations and organizations. The list of sponsors has recently expanded to include governments outside the United States. Total funding has increased from approximately $150,000 in 1972 to almost $3,000,000 in the fiscal year 1978-1979.

RESEARCH INTERESTS:

Each senior SSRI scientist is encouraged to pursue his or her own research interests, subject to availability of funding. These interests are diverse; a recent count identified 27. Four major interests persist among groups of SSRI researchers: crime control and criminal justice, methods of dispute resolution and alternatives to the courts use of administration records for demographic and other research purposes, and exploitation of applications of decision analysis to public decision making and program evaluation. But many SSRI projects do not fall into these categories. Most projects combine the skills of several scientists, often from different disciplines. As SSRI research personnel change, its interests will change also.
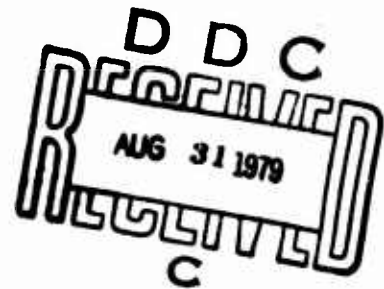
VALIDATION AND ERROR IN MULTIPLICATIVE

UTILITY FUNCTIONS

F. Hutton Barron

Social Science Research Institute
University of Southern California

DDC
RECEIVED
AUG 31 1979
C

December, 1978

79 08 31 058

SUMMARY

In this report an approach to the concept of error in utility assessment is proposed. Three components of error are considered and each component is related to four separate elicitation methods -- all in the context of a general multiplicative multiattribute utility model. The methods are a Keeney-Raiffa (1976) procedure, SMART (Edwards, 1977), a social judgment theory (SJT) based regression model (Hammond, Stewart, Brehmer and Steinmann, 1975) and a new method called Holistic Orthogonal Parameter Estimation or HOPE (Barron and Person, 1978).

If a general multiplicative model can be assumed to be an appropriate representation of the decision maker's basic preference structure, error can occur in the direct estimation of the scaling constants and univariate utility functions for decomposition methods (Keeney-Raiffa and SMART), or in the holistic assessments for holistic methods (SJT and HOPE). Individual estimates may be merely noisy or may be fundamentally incorrect. Furthermore, the utility model may be incorrectly specified; for example, an additive model, rather than a multiplicitive model, may be used. The four assessment methods are considered in conjunction with errors of each kind.

The most serious error-method combination is the case of a substantial degree of error occurring in a single holistic judgment which is being used in a HOPE procedure. This concern leads to a major emphasis of this report -- an expanded HOPE procedure used in conjunction with a convergent validation strategy to estimate error in individual holistic judgments and thus guide consistency checks.

The discussion is organized into four sections. The HOPE procedure is summarized in Section I. In Section II, three components of assessment error are considered in conjunction with the four elicitation procedures. In Section III, an expanded HOPE procedure for detecting judgment error and guiding consistency checks is proposed. In Section IV, application considerations are outlined.

CONTENTS

Appendix

Tables

ACKNOWLEDGEMENT

Disclaimer

The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or of the United States Government.

Introduction

In this report an approach to the concept of error in utility assessment is proposed. Three components of error are distinguished; these components are then related to four separate elicitation methods -- each of which is consistent with at least special cases of the general multiplicative multiattribute utility (MAU) model (below). Two methods, Keeney-Raiffa (1976) and SMART (Edwards, 1977) are pure decomposition approaches; a third, the social judgment paradigm (Hammond, Stewart, Brehmer, and Steinmann, 1975) is a regression approach which relies on holistic judgments.

The fourth approach is a decomposition procedure for assessing multiplicative MAU functions which relies solely on a few holistic assessments of utilities. The procedure's acronym is HOPE for Holistic Orthogonal Parameter Estimation. Consistent with the procedures of Keeney and Raiffa (1976), the HOPE procedure exploits the basic preferences of the decision maker to specify the utility function. HOPE differs in that it uses a response mode more familiar to laymen than those of other methods of MAU elicitation -- holistic assessment of (a few) profiles -- to determine the scaling constants and univariate utility functions comprising the multiplicative utility function.

The larger question behind any analysis of error in assessed utility functions concerns validation. There are three basic approaches to validation of assessed utility functions: (1) use of an external criterion; (2) validating the basic preference structure of the decision maker; (3) convergent validity. Of these, the most straightforward is use of an objective, externally defined, criterion, if one is available.

Edwards (1974) has suggested two instances of available criteria --
diamonds and bank credit. The American Gemological Institute "diamond
model" formally evaluates diamonds based on the four attributes: color,
cut, clarity, and carats. Banks evaluate applicants for credit cards on
the basis of attributes contained on standard application forms (e.g.,
disposable income, own versus rent, debt, employment history, etc.);
while actual probability and amount of default, if any, are known
empirically.

Edwards and his associates have suggested a variant of the multiple
cue probability learning paradigm (Hammond, Stewart, Brehmer and Stein-
mann, 1975) as a means of creating an external criterion. Subjects are
first trained to use a weighted additive utility functions; this is
followed by eliciting learned utilities. This procedure has been pilot-
tested at SSRI.

The purpose of studies using an external criterion is to compare
and evaluate the performance of elicitation techniques. The usefulness
of such studies rests on the assumption that comparative efficacy of
elicitation methods generalizes from situations where an external
criterion exists to situations where one does not. Of course, external
criteria do not generally exist in de novo assessment of utilities in
laboratory or field settings. Furthermore, elicitation is unnecessary
when such a criterion exists. Thus, in application as opposed to ex-
perimental settings, validity necessarily depends on either validation
of the basic preference model itself or on convergent validity.

Validating the basic preference model requires verifying qualitative properties of preferences which are sufficient for representing preferences by a particular mathematical model. In the decomposition approach of Keeney and Raiffa (1976), the qualitative properties of preferential independence and utility independence are explicitly examined. Preferential independence requires that preference orders over attribute pairs be independent of the fixed levels of the remaining attributes. Utility independence requires that preferences for lotteries involving a single attribute be independent of the fixed levels of the remaining attributes. If these qualitative properties of preferences hold, then the preference structure can be represented by either a multiplicative model, equation (1), or by an additive model, equation (2).

$$1 + KU(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} (1 + Kk_i u_i(x_i)) \tag{1}$$

where, for i = 1, 2, ..., n

   x_i is level i of attribute X_i

   $0 \leq u_i(\cdot) \leq 1$ is a utility function

   $0 \leq k_i \leq 1$ is a scaling constant

   $-1 \leq K$ is a parameter

   $U(\cdot)$ is overall utility.

Upon writing equation (1) simply as an expression for $U(\cdot)$ and taking limits as K goes to 0, the resulting model is the additive model, equation (2):

$$U(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} k_i u_i(x_i) \tag{2}$$

The basic preference model is validated by checking the conditions of preferential and utility independence. For an especially informative assessment protocol in which the conditions are checked, see Keeney (1977).

The choice between models 1 and 2 in an application setting is not necessarily a simple choice. The choice can be viewed as an empirical question, in which the parameter $K = 0$ if and only if $\sum_{i=1}^{n} k_i = 1$. The choice can be based behaviorally on the marginality condition (Fishburn, 1965). Model 2 is appropriate only if the decision maker's preferences for multiattributed risk consequences depend only on the marginal distribution associated with each attribute. Experimental studies (e.g., both conducted and surveyed by von Winterfeldt (1976)) have often found violations of the marginality condition. Finally, the choice can be made on more pragmatic grounds. Additive models are simpler, provide excellent predictions, and can be used in conjunction with simpler procedures (e.g., the SMART procedure discussed in this report). In fact, referring to the same experimental studies von Winterfeldt states:

> The message that these experiments convey seems contradictory: in spite of obvious model violations (tests of marginality and tests of risky additivity failed), additive models . . . predict subjects' preferences and utility judgments very well (p. 24).

Convergent validation assumes that logically equivalent elicitation procedures will assign comparable utilities to the same multi-attributed outcome. In his convergent validation study Fischer (1977) elicited utilities for 27 hypothetical jobs from each of 10 subjects in three different ways: (1) direct holistic assessments, (2) via a Keeney-Raiffa decomposition, and (3) via the SMART procedure. The utilities

predicted for the 27 jobs (within subject) by the Keeney-Raiffa and SMART procedures exhibited high correlations with each other and were highly correlated with the holistic assessments.

Slovic, Fischhoff and Lichtenstein (1977) have criticized using correlation between predictions of MAU models and holistic judgments as evidence that the MAU model is valid. Their criticism is based on the use of unaided holistic preference as a criterion. They further state, ". . . a decade or more of research has abundantly documented that humans are quite bad at making complex unaided decisions (Slovic [1972]); it could thus be argued that high correlations with such flawed judgments would suggest a lack of validity" [p. 22].

At face value the remarks of Slovic et al. are damaging to the HOPE procedure described in this report. In fact, a major emphasis of this report is the use of an expanded HOPE procedure in conjunction with a convergent validation strategy to estimate prediction (of utilities) error, and to thus identify outliers in the set of holistic judgments. Thus, the view represented by the remarks of Slovic et al. must be considered.

I believe decision makers can provide useful holistic assessments of (a few) multiattributed consequences, especially in cases involving a small number of attributes, say fewer than ten. Furthermore, there is considerable evidence to support my belief. Two research traditions, heavily based on holistic assessment, are social judgment theory (Hammond et al.) and information integration theory (Anderson [1974]). Numerous empirical studies in these traditions alone constitute examples of

judges providing holistic judgments. Of course, neither the cognitive processes which produce holistic judgments are well understood, nor are the exact parameters of such judgments known. It is generally accepted that such judgments are often systematic, but noisy (Fischer, 1977); in fact, bootstrapping (Dawes and Corrigan, 1974) capitalizes on these features. It is also presumed that noise increases as the number of attributes increases. Finally, it is plausible to presume that noise increases as the number of required holistic evaluations increases.

The HOPE procedure requires direct holistic assessment of the utility of each of a small number of multiattributed consequences comprising a highly fractionated experimental design. By small number, I mean fewer than fifty consequences, a number suggested by both Keeney and Raiffa (1976, p. 222) and Fischer (1977). My basic assumption is that decision makers can provide the requisite holistic assessments; furthermore I recognize that these assessments will be noisy. Moderate noise, per se, does not materially reduce the efficacy of the HOPE procedure. A previous paper (Barron and Person, 1978) demonstrated that the HOPE procedure could recover known MAU functions from simulated noisy holistic judgments. Recovery was excellent in the presence of moderate amounts of noise -- defined as normally distributed additive error having a standard deviation of .05. Model specification error -- defined as the use of an additive model fitted to error-free holistic judgments computed from known multiplicative models -- produced higher prediction error than did moderate noise in conjunction with an appropriate multiplicative model using an estimated K value.

The HOPE procedure generalizes a procedure for estimating main effects in an <u>additive</u> model (Green, 1971), to the assessment of the <u>multiplicative</u> utility functions of Keeney (1974). Multiplicative (and additive) utility functions are of interest for three reasons. First, the multiplicative model represents the decision maker's preference structure if certain preferential and utility independence conditions are satisfied. In those cases in which these conditions have been verified, the multiplicative model is valid. Second, the multiplicative utility function is of practical importance, even if the requisite assumptions do not hold precisely (Keeney and Raiffa, 1976, p. 298). Third, formally the general multiplicative model encompasses the kinds of utility functions resulting from each of the four assessment approaches considered in this paper.

This report is organized into four sections. The HOPE elicitation procedure is described in section I. In section II, three components of utility assessment error are considered in conjunction with the four elicitation procedures. In section III, a method for detection of judgment error is proposed and illustrated by application to the data of Fischer (1977). In section IV, application considerations are outlined.

I. HOPE: A Utility Elicitation Procedure[†]

The HOPE and Keeney-Raiffa[*] procedures differ only in the parameter estimation phase. HOPE estimates the parameters -- the univariate

---

[†]Section I is a revised version of section I of Barron and Person (1978).

[*]Unless noted otherwise, by "a Keeney-Raiffa procedure" is meant their general approach to assessing a multiplicative utility function as described in Keeney and Raiffa, 1976, pp. 297-304.

utility functions and scaling constants -- of the multiplicative family of utility functions from holistic judgments of utility. The estimates are derived; they are inferences which are completely based on the appropriateness of the underlying preference structure. A Keeney-Raiffa procedure, while relying on the appropriateness of the same underlying preference structure, differs, primarily in that scaling constants and utility functions are individually assessed.

The HOPE procedure consists of three phases: (1) preparation, meaning those aspects common to both HOPE and Keeney-Raiffa approaches (necessary preliminaries, identification and definition of attributes, determination of value ranges, verification of appropriate independence conditions); (2) eliciting direct holistic assessments for the specific multiattributed consequences of an appropriate orthogonal design; (3) performing the arithmetic on the holistic assessments required to deduce scaling constants and utility functions.

Preparation

All approaches to multiattribute utility assessment involve certain necessary preliminaries. Clearly the stage must be set, the respondent must realize the purpose of the exercise, and mutual understanding sufficient for effective communication must be established. Following preliminaries, both approaches would identify and define value-relevant attributes. For each attribute, both would determine an appropriate range of values, including which level is worst and which is best. Finally, both would check to see that necessary preferential independence

and utility independence assumptions are met.[*] As a consequence of the independence conditions, both would conclude that Keeney's general multiplicative model, equation (1), is an appropriate MAU function. At this point, a Keeney-Raiffa procedure differs from the HOPE procedure by assessing utility functions and scaling constants individually. The utility functions, u_i, are assessed via standard gambles in the usual way (Keeney and Raiffa, Chap. 4). The parameter k_i is interpreted and often in practice assessed as an indifference probability in the standard gamble which yields consequence BEST (defined as all attributes at their best levels) with probability k_i and consequence WORST (all attributes at their worst levels) with probability $1 - k_i$, versus the certain consequence with attribute X_i at its best level and all other attributes at their worst levels. Scaling constants can also be assessed in other ways. In the HOPE procedure, neither the utility functions, u_i, nor the scaling constants, k_i, are directly assessed. Rather, the HOPE procedure infers utility functions and scaling constants from holistic assessments of consequences defined by an appropriate orthogonal design as described below.

Orthogonal Arrays Define Consequences for Holistic Assessment

Orthogonal experimental designs generally require the lowest number of holistic consequence assessments for additive main-effect non-confounded parameter estimation. A catalog of useful orthogonal designs and variations is provided by Addelman (1962).

---

[*]The qualitative properties of preferential and utility independence are described succinctly in Keeney (1977, p. 271) and amplified in Keeney and Raiffa (1976, Chap. 6).

An example of a particular design appropriate for up to 5 attributes with up to 4 levels per attribute is Addelman's "Basic Plan 3" shown in Table 1. For example, consequence 3 in Table 1 is defined as level 1 of attribute 1, level 3 of attributes 2 and 3, level 4 of attribute 4, and level 2 of attribute 5. This consequence, along with the 15 others of Table 1, plus one additional consequence defined in Appendix 1 and used to estimate the parameter K, would then be holistically evaluated. Note that consequence 1 represents the worst level of each attribute, and would be assessed to have zero utility in the preparation phase. A worst reference outcome is common to the orthogonal designs on which HOPE is based.

Holistic responses may be either direct ratings, appropriate for riskless utilities, or standard gamble indifference probabilities, appropriate for risky utilities. For riskless utility assessment one reference case consisting of the best level of all attributes is as-signed 100 points; a second reference case defined as the worst level of all attributes is assigned 0 points. The remaining consequences defined by the experimental design are then rated individually along the 0 to 100 point scale. Rating data, normalized by dividing by 100, are then treated as if they were interval-scaled responses. For lottery-type utility assessments the design consequences could be considered as sure-thing consequences in a standard gamble with the same two reference cases as uncertain outcomes. Related riskless procedures using 6-to-11

point direct rating scales have been used in marketing applications[*]
such as hospital promotion (Wind and Spitz, 1976).

Analysis (Arithmetic) of Holistic Responses

Succinctly stated, the HOPE procedure receives m noisy holistic
judgments as input. Using 2 judgments, the parameter K is estimated
(step 1, Appendix 2). Depending on the value of K, either m - 1 judg-
ments are subjected to the arithmetic of the additive model (steps 2-5
of Appendix 2), or m - 1 transformed judgments are subjected to the
arithmetic of the multiplicative model (steps 6-10 of Appendix 2). The
result or output is a complete set of k_i and $u_i(x_i)$ values, the latter
for each level of each attribute specified by the design.

The analysis of the holistic evaluations is made extremely simple
by the use of an orthogonal design. It merely involves arithmetic. The
main effect of any given level of any given attribute is then the sum of
the holistic values of all consequences containing the attribute at the
given level (divided by the number of such consequences), minus the
corresponding sum of the holistic values of all consequences containing
the worst level of the same given attribute (again divided by the number
of such consequences). If the procedure described in Appendix 1 yields
an estimated K = 0, then the additive model, equation (2), is appro-
priate, and the estimated main effects must be normalized by dividing
each estimate by the sum of the estimated best levels. For each attri-
bute, this procedure estimates $k_i u_i(x_i)$ in general and k_i when X_i is at

---

[*] Carmone, Green and Jain (1977) cite a figure of over 200 indus-
trial applications of additive (conjoint measurement) models ranging
from use of full factorials and ranking responses to use of orthogonal
arrays and rating responses. These applications assume equation (2) to
be an appropriate model.

its best level since $0 \leq u_i(x_i) \leq 1$. Otherwise, the multiplicative model, equation (1), is appropriate. In this case each holistic value, U, is first transformed to $\ln(1 + KU)$, where ln is the natural logarithm. The above analysis without normalization is then simply carried out on the transformed holistic values. For the multiplicative model this procedure estimates the quantity $\ln(1 + Kk_i u_i(x_i))$ in general, and $\ln(1 + Kk_i)$ when X_i is at its best level. Since K is estimated separately it is then possible to compute all k_i and $u_i(x_i)$ values.

II. Error in Assessed Utility Functions

Assuming the general multiplicative model _is_ an appropriate representation of the basic preference structure, error can occur in the direct estimation of the scaling constants and utility functions for decomposition methods or in the holistic assessments for holistic methods. The individual estimates may be merely noisy, or may be fundamentally incorrect. In predicting utilities, the analyst may further mis-specify the model (e.g., may choose equation (2) rather than equation (1)). Thus, prediction errors may be related to one or more of the following errors: (1) model specification error; (2) noisy subjective estimates; (3) substantial random error.

It is difficult to characterize substantial errors of judgment. Surely, a substantial error is a judgment which would be altered upon reflection. It would exhibit low test-retest reliability. In a statistical sense a substantial error is an outlier.

Examples of the three types of error occurred in my recent field study of professional audit judgments. In that study audit partners

considered hypothetical client firms described by sets of financial statements. For each firm a dollar amount defining material error was estimated to guide statistical sampling procedures. Some dollar amounts were used to estimate the parameters of a judgment model; the remaining estimates constituted a holdout sample to be predicted from the model. Several sources or error could have contributed to prediction error. First the assumed additive prediction model could have been an incorrect specification. Second, the original estimates were stated to the nearest $25,000. With this type of rounding, a "true" value of say, $190,000 would sometimes have been assessed as $175,000 and sometimes as $200,000. Or sometimes, the estimate would have been "either $175,000 or $200,000." Third, any of the judgments to be predicted or the judgments used to estimate the parameters could have been either merely noisy or substantially wrong. Test-retest cases and deviations from additive predictions indicated the general level of noise. In instances where there was substantial prediction error, participants were asked to reexamine cases whose estimated dollar amounts either influenced parameter estimates or represented holdout cases. For many firms the original dollar amount was deemed "correct," but in a few cases judged dollar amounts were substantially revised. Comments like "I don't know what I could have been thinking," accompanied such revisions. Yet clearly there was no definitive criterion for "noise" versus "substantial error."

In the remainder of this section we examine the four elicitation procedures, designated KR (for Keeney-Raiffa), SMART, SJT (for social judgment theory) and HOPE, in conjunction with the possible effect on each of specification error, noise, and substantial random error.

Model Specification Error

All four methods estimate univariate utility functions and scaling constants for either the additive model equation (2), or the multiplicative model equation (1). SMART and SJT consider only additive models. Thus, a priori, SMART and SJT models are incorrectly specified for all true values of the multiplicative parameter $K \neq 0$.

There are several studies which indicate the effect of specification error within elicitation methods. Fischer (1977) observed high correlations (median = .982) between KR additive predictions and KR multiplicative conditions. Simulating the HOPE procedure using known multiplicative models with extreme K values (K = -.94 or K = 4.11) with two levels of response noise (normally distributed, mean 0 and standard deviations of .025 and .05), Barron and Person (1978) found larger errors of predictions for incorrectly specified additive models (with or without noise) than for correctly specified models with noise. Furthermore, as the standard deviation of response error was decreased from .05 to .025 prediction error for correctly specified multiplicative models was cut in half, while for incorrectly specified additive models with noise, prediction error persisted at 80% and 90% of its former level. Analysis of Fischer's risky data by HOPE procedures shows for 8 of 10 subjects multiplicative models produce a lower root mean square error in predicting the entire set of holistic judgments than do additive models. A statistical analysis of the same risky data by Fischer found 6 of 10 subjects departed significantly (α = .05) from additivity and for each of the 6 cases, multiplicative models produced a lower standard error of

estimate. Finally, all 30 estimates of K (3 estimates for each of Fischer's 10 risky data sets) using the procedure described in Appendix 1 were found to be different from 0.

Analysis of the laboratory data of Fischer and the simulation data of Barron and Person suggest that model specification error increases prediction error. The HOPE method provides a simple direct estimate of K. The KR method computes K as a function of the scaling constants, while SMART and SJT work directly with relative weights assuming $K = 0$.

The individual methods differ in their respective approaches to specification of the univariate utility functions, $u_i(\cdot)$. KR and SMART assess the u_i directly. SMART asks respondents to assign $u_i(x_j)$ values directly to selected levels x_j of attribute i, $i = 1, 2, \ldots, n$. KR assess a univariate utility via standard gambles -- finding a few certain equivalents, followed by fairing in a curve. Direct assessment of certain equivalents is especially vulnerable to Tversky's (1977) suggested certainty effect bias.

The HOPE and SJT methods infer the u_i statistically from the holistic responses. SJT uses multiple linear regression; candidate u_i functions are polynomial functions of x_j. The ability of linear models to account for substantial proportions of variance (e.g., Yntema and Torgerson, 1961; Slovic and Lichtenstein, 1971) often leads to u_i functions which are linear in x_i. The HOPE procedure provides a single point estimate for $u_i(x_i)$ for each x_i specified by the design. Since there are no degrees of freedom for error estimation in the single design HOPE procedure, a substantial error in a single holistic judgment could produce erroneous u_i functions.

With respect to weights, and more generally, scaling constants, SMART directly assesses relative weights which are then normalized; KR assesses scaling constants directly via standard gambles or sets of equations reflecting specific tradeoffs (Keeney and Raiffa, p. 267); HOPE infers scaling constants consistent with the ex ante model specification; and SJT infers beta weights from the regression. SJT weights are influenced by both model specification error (additive only) and univariate utility function specification error (usually linear).

Tradeoffs between judgment error and modeling error are complex. SMART weights are simpler to assess than are KR scaling constants. Simpler judgments (relative weights) may outweigh the disadvantage of model specification error (assuming $K = 0$). By contrast, SJT and HOPE procedures use equivalent holistic judgments. In this case assuming an additive model, i.e., $K = 0$, serves as a constraint.

Noise

KR and SMART procedures estimate scaling constants or weights and univariate utility functions separately. Thus noise in the estimate of one set of parameters should not effect estimates of the other set. Of course, the values of the scaling constants depend on the attribute ranges. SMART weights are normalized by the sum of the estimated values. Thus SMART estimates are sensitive only to errors in the relative values of these estimates. Keeney (1977, p. 284 ff) first ranks attribute (ranges) by importance and then assesses specific values via standard gambles and/or tradeoffs. Standard gambles will underestimate scaling constants to the extent they are subject to Tversky's certainty effect.

Using direct tradeoffs, noise in the univariate utility function values will lead to noise in scaling constant values.

SJT infers the scaling constants assuming linear utility functions via regressing standardized attribute levels against holistic judgments. The assumption of linear conditional utility functions introduces error; a second source of error is noise in the holistic judgments. The usual error theory of regression analysis provides estimates of the sensitivity of the inferred scaling constants (beta weights).

HOPE simultaneously infers conditional utility functions and scaling constants from holistic judgments. The orthogonal arrays utilized by this procedure are extremely efficient regression designs. There are, in fact, zero degrees of freedom. If the value of a single holistic judgment is changed, the estimated value of at least one scaling constant, and several u_i values will change. If the holistic judgment is merely noisy, the effect on k_i and $u_i(\cdot)$ are minimal; if the judgment reflects "substantial random error," the effects will be substantial. Each of these last two statements finds support in the individual simulation runs of Barron and Person (1978).

Substantial Random Error

"Holistic evaluative judgments are characterized by a substantial degree of random error," states Fischer (1977, p. 296). If so, the SJT, HOPE, and KR assessments are affected. SJT and HOPE use only holistic assessments as data; KR assessments of scaling constants in a standard gamble context require a holistic assessment of outcomes having one

attribute at its best level and all other attributes at their worst levels. In this latter case, Tversky's certain effect bias is confounded with the predicted "substantial degree of random error."

Judgment error per se is not a part of the formal theory of MAU. Error may be handled implicitly via sensitivity analysis. In the assessment stage, the careful analyst performs numerous consistency checks in an attempt to prevent substantial error. For example, in SMART, if the relative values of attribute ranges I, II, and III are 10, 30, and 60 respectively, i.e., II is 3 times as important as I, while III is 6 times as important as I; then as a consistency check, III should be judged 2 times as important as II. For KR procedures, numerous consistency checks are illustrated in a detailed protocol (Keeney, 1977).

The most serious error suggested thus far is a single holistic judgment exhibiting a substantial degree of error being used in a HOPE estimation procedure. Since there are zero degrees of freedom, the error affects several k_i or u_i values individually. Since the error is substantial the impact, though moderated by other judgments, is also substantial. Thus, it is essential to identify, if possible, individual holistic judgments exhibiting substantial random error.

In the next section a modified HOPE procedure is described. The modified procedure guides consistency checks over the original set of holistic judgments and provides a means for detecting substantial random error.

III. Detection of Error in Holistic Judgments

The HOPE procedure can be easily extended to detect substantial random error over the set of holistic judgments. The extended HOPE procedure uses two orthogonal designs with minimal overlap in one of two possible ways. First, the data of both designs is pooled to estimate k_i and $u_i(x_i)$ values. This serves to increase the degrees of freedom in the HOPE procedure to approximately the number of design points unique to one design. Noise is then defined by the root mean square error, i.e., of pooled prediction minus actual judgments, while arbitrarily, substantial error is defined to be a deviation exceeding twice the root mean square error.

A second approach using two designs is to build two separate (utility) prediction models -- one for each design. The utility pre-dictions of design 1 (2) are used to predict the actual judgments of design 2 (1). Substantial error is again arbitrarily defined by pre-diction errors exceeding twice the root mean square error.

Either of the above procedures can be used to "detect" excessive deviations from model predictions, although it may be more reasonable to consider these procedures as guides for checking consistency over the original set of holistic judgments. One set of judgments which are candidates for consistency checks are those judgments which deviate substantially from predicted values. A second set of candidates are equal judgments for which the predicted values diverge. For example, if two consequences, a and b, have judged utilities $u(a) = u(b) = .6$ but the predictions differ; say $u'(a) = .64$ and $u'(b) = .57$; then a and b may also be presented to the subject for reconsideration.

Each error detection procedure can be illustrated using the data of Fischer (1977). Fischer's subjects provided holistic evaluations for 27 hypothetical offers of employment. Each job was (completely) described by 3 attributes -- salary, location (city), and type of work. Each attribute had 3 possible levels, designated in the tables and discussion as W, for worst level; I, for intermediate level; and B, for best level. The 27 jobs represent all combinations of each level of each attribute. Fischer (1977) provides a detailed description of the elicitation task. These data have been analyzed via conjoint measurement (Fischer, 1976), have been subjected to convergent validation tests in conjunction with several elicitation methods (Fischer, 1977), and have been subjected to various HOPE procedures (Barron, 1978). For illustrative purposes, we consider the (risky) responses of subject 2, a subject for whom the additivity hopotheses (i.e., equations 1 and 2) were rejected by conjoint measurement analysis (Fischer, 1976, p. 139).

A double design appropriate for the analysis of subject 2's responses is presented in Table 2. Predicted values for each design are based on pooling data from both designs. Design 1 indicates two points for consistency checks (I,W,I) and (B,I,W). Point (I,W,I) is also a design 2 point; except for (I,W,I), design 2 has no points indicated for consistency checks. These two points should now be reconsidered by the person making the original judgments. (Of course, using these data supplied by Fischer, this is impossible.)

At this point, let us assume that upon reconsideration subject 2 agreed to reduce the judged (B,I,W) from .6l' to .50. If this one change

is made, and the predictive model for the revised pooled data is obtained,
several things happen as shown in Table 3. The new predictions for
(B,I,W) and (I,W,I) are both brought into better agreement with stated
values -- (B,I,W) because the stated value was revised and (I.W,I) be-
cause the predicted value based on the revised (B,I,W) value increased.
All values are brought into better agreement as measured by root mean
square error (RMSE). Revised RMSE over all 16 pairs is smaller than the
prior RMSE either including or excluding the (B,I,W) pair.

Consistency checks should also be performed for sets of equal judg-
ments as guided by model predictions. In design 2 cells (B,I,I) and
(B,B,W) were each assigned a value of .65. The predicted values are .70
for (B,I,I) and .59 for (B,B,W). The subject should be asked to recon-
sider these judgments. Is (B,I,I) really preferred to (B,B,W)? A
similar comment does not apply to design 2 cells (W,I,B) and (W,B,I),
each assigned original values of .70. The predicted values differ, .66
and .69 respectively, but do not diverge.

Using the two designs to estimate separate predictive models pro-
duces similar results. Cells (I,W,I) and (B,I,W) of design 1 are iden-
tified for consistency checks. Revising the (B,I,W) utility suggests
the original (I,W,I) value is not substantially wrong. Furthermore,
design 1 predictions of the equal value cells (B,I,I) and (B,B,W) sug-
gest the former's utility should be increased and the latter's should be
decreased.

The use of double designs provides a mechanism for detecting sub-
stantial error in judgment and further guiding consistency checks. The

error detection feature comes at the cost of requiring approximately twice as many judgments as originally required. Its advantage is that it provides a within-elicitation method for convergent validation of field assessed utility judgments.

IV. Application Considerations

Proponents of each elicitation method described can point to an impressive number of applications. Illustrative of KR are Keeney (1973, 1976, 1977); of SMART, Edwards (1977), Gardiner and Edwards (1975); of SJT, Hammond and Adelman (1976) and of HOPE, Wind and Spitz (1976), and those cited in Carmone et al. (1977). This section will provide neither an extensive analysis of each method nor an applications critique of each method. Instead, practical considerations will be highlighted.

Keeney has reported two applications, Keeney (1976, 1977) that provide special insight into the process of elicitation itself -- a process which is intensive, demanding and dynamic. In each case, respondents are highly motivated professionals who had thought deeply about the respective problems, and Keeney is an especially skillful assessor. Generalizing from these reported applications, a KR procedure is expensive in terms of both respondent and assessor time. It is also an intensive process requiring a rather skilled assessor. Respondents are often required to be conversant with probability concepts. There is no estimate of "noise," and no particular procedure guides consistency checks. When the assessment is finished, the result is a utility function. Random error may implicitly be considered via sensitivity analyses.

The procedure's advantages are its obvious tie to underlying theory, its attention to model specification, and the likely independence of noise in the scaling constants and univariate utility functions. Its disadvantages are the requirements that possibly unfamiliar constructs (scaling constants and univariate utility functions) must be directly assessed in a possibly unfamiliar language (probability theory).

The primary advantage of the SMART procedure is its simplicity. Based on simple rating procedures to deduce weights and utility functions, it has the further advantage of being easy to teach to (probabilistically) naive decision makers. It is easily adapted to hierarchical utility structures. Although the procedure is believed to be robust, a disadvantage is its sole reliance on the additive model. An error detection procedure for relative weights consists of a triangular matrix of ratios as described earlier in this report.

SJT and HOPE procedures rely on holistic judgments. It is generally acknowledged that people are capable of assigning values holistically in a consistent and meaningful fashion provided the number of attributes is, say, less than 10 and the total number of evaluations is, say, less than 50, although occasional judgments are subject to substantial error. Thus, it can be argued that each procedure uses psychologically meaningful and familiar judgments. The disadvantages of SJT include reliance on the additive model, equation (2), and a tendency to infer linear utility functions. There is also the difficulty of detecting error in individual judgments.

HOPE has been characterized as a "decomposition procedure which relies solely on holistic assessments" (Fischer, personal communication). It includes procedures designed to both aid correct model specification and detect substantial error in individual judgments. Its obvious limitations are that it requires credible orthogonal attribute combinations and it does not apply to value hierarchies.

The HOPE procedure seems to have certain advantages when probing utilities of large numbers of people. For example, first a few representative individuals can be subjected to an intensive process. Here the reasonableness of preferential and utility independence properties can be checked and unimportant attributes can be discarded. Next, a larger sample of respondents can be assessed via questionnaire. Subsequently questionnaire assessments could be checked through a small sample of follow-up interviews.

The HOPE procedure is an extension of an additive conjoint measurement approach to modeling consumer preferences for multi-attribute alternatives (Green and Rao, 1971). Early applications of the conjoint methodology, e.g., Green, Carmone, and Wind (1972), relied on full factorials, rank-order responses, and nonmetric scaling procedures. Green, et al., have subsequently simplified the basic approach considerably. Orthogonal designs significantly reduce the number of consequences to be evaluated from a prohibitive to a feasible number. Direct rating of a few consequences is often perceived as less tedious than ranking. Metric recovery requires substantially less computational capacity.

The efficacy of these simplifications has been supported in the recent simulation studies of Carmone, et al. (1977), and Barron and Person (1978). That there have been over 200 industrial applications of conjoint analysis (using both full factorials and, more recently, orthogonal arrays) attest to HOPE's practicability. The refinements of correct model specifications and error detection can only enhance its usefulness.

Having practical alternatives to KR and SMART provides obvious advantages. Assessment can be more easily tailored to the specific situation -- considering costs, nature of respondents, and importance of the contemplated decision. The particular strengths and weaknesses of different procedures can be further determined through actual practice and empirical research.

Appendix 1: Estimation of the Parameter K

Consequence 13 (Table 1) denoted below by "a," consists of the highest levels of attributes 1, 3, 4, 5 and the lowest level of attribute 2. If one additional consequence, denoted "b," is defined to have attribute 2 at its highest level and attributes 1, 3, 4, 5 at their lowest levels and is included with the orthogonal design, then ratings of these complementary consequences a and b can be used to estimate K.

If the MAU is additive, then the sum of the observed utilities for the complementary consequences is 1.0. Otherwise define consequence "c" to have all attributes at their highest levels. By regrouping terms from the right side of equation (1), we have the following:

$$1 + k\,u(c) = (1 + Kk_2) \cdot (1 + K\,u(a)) \tag{3}$$

$$k_2 = u(b) \tag{4}$$

$$u(c) = 1 \tag{5}$$

Substituting (4) and (5) into (3) and solving gives

$$K = (1 - u(a) - u(b))/(u(a)\,u(b)).$$

In practice, the ratings u(a) and u(b) are estimates, so K is also an estimate.

Appendix 2: Illustration of the Arithmetic
of the HOPE Procedure

Assume the decision maker has provided holistic assessments for the 16 consequences of Table 1. Designate these values h_1, h_2, ..., h_{16}. Designate by h_{17} the assessment of consequence (1, 4, 1, 1, 1), that is, the consequence having attribute 2 at its best level and all other attributes at their worst levels. The arithmetic proceeds as follows:

Step 1. Following Appendix 1, compute $h_{13} + h_{17}$. If $h_{13} + h_{17} = 1$, then $K = 0$ and we estimate the parameters of the additive model (steps 2-5). If $h_{13} + h_{17} \neq 1$, proceed to step 6 to estimate the parameters of the multiplicative model.

Step 2. $A_1 = (h_1 + h_2 + h_3 + h_4)/4$
$A_2 = (h_5 + h_6 + h_7 + h_8)/4$
$A_3 = (h_9 + h_{10} + h_{11} + h_{12})/4$
$A_4 = (h_{13} + h_{14} + h_{15} + h_{16})/4$

Step 3. $k_1 = A_4 - A_1$
$k_1 u_1$ (level 3 of x_1) $= A_3 - A_1$
$k_1 u_1$ (level 2 of x_1) $= A_2 - A_1$

Step 4. Repeat steps 2 and 3 for each attribute. Note the definition of A_1, A_2, A_3, A_4 changes for each attribute, as defined by the design. For example, for attribute 2
$A_1 = (h_1 + h_5 + h_9 + h_{13})/4$
$A_2 = (h_2 + h_6 + h_{10} + h_{14})/4$ etc.

Step 5. Compute the sum $k_1 + k_2 + \ldots + k_5$. If this sum does not equal one, then normalize all values computed in step 3 for each attribute by dividing each value by this sum.

Step 5 completes the estimation process for the additive model. Steps 6-10 describe the estimation process for a multiplicative model.

Step 6. First calculate K', an estimate of K, using the following relationship derived in Appendix 1:

$$K' = (1 - h_{13} - h_{17})/h_{13}h_{17})$$

Step 7. Define $h_i' = \ln(1 + K'h_i)$, $i = 1, 2, \ldots, 16$

Step 8. Repeat step 2 (additive model) using h_i' values.

Step 9. $1 + K'k_1 = e^{(A_4 - A_1)}$

$1 + K'k_1u_1$ (level 3 of x_1) $= e^{(A_3 - A_1)}$

$1 + K'K_1u_1$ (level 2 of x_1) $= e^{A_2 - A_1)}$

Step 10. Repeat steps 8 and 9 for each attribute. As before, the definition of A_1, A_2, A_3, A_4 are given by the design and <u>differ</u> for each attribute.

References

Addelman, S. Orthogonal main effect plans for asymmetrical factorial experiments. Technometrics, 1962, 4, 21-46.

Anderson, N. H. Information integration theory: a brief survey, in D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes (Eds.). Contemporary Developments in Mathematical Psychology, Vol. 2, San Francisco: Freeman, 1974.

Barron, H. F. & Person, H. B. Assessment of multiplicative utility functions via holistic judgments. Bayesian Research Conference Paper, Los Angeles, 1978.

Carmone, F. J., Green, P. E., & Jain, A. K. The robustness of conjoint analysis: some Monte Carlo results. Journal of Marketing Research, 1978, XV, 300-303.

Dawes, R. M. & Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.

Edwards, W. Social choice in a multiattribute utility environment. AIDS Decision Analysis Workshop Paper, Atlanta, 1974.

Edwards, W. Use of multiattribute utility measurement for social decision making in D. E. Bell, R. L. Keeney, & H. Raiffa (Eds.), Conflicting objectives in decisions, New York: Wiley, 1975.

Fischer, G. W. Multidimensional utility models for risky and riskless decisions. Organizational Behavior and Human Performance, 1977, 18, 295-315.

Fischer, G. W. Convergent validation of decomposed multiattribute utility assessment procedures for risky and riskless choice. Organizational Behavior and Human Performance, 1976, 17, 127-146.

Fishbein, P. Independence in utility theory with whole product sets. Operations Research, 1965, 13, 28-45.

Gardiner, P. & Edwards, W. Public values: multiattribute utility measurement for social decision making, in M. Kaplan and S. Schwartz (Eds.), Human Judgment and Decision Processes, New York: Academic Press, 1975.

Green, P. E., Carmone, F. J., & Wind, Y. Subjective evaluation models and conjoint measurement. Behavioral Science, 1972, 17, 288-299.

Green, P. E. & Rao, V. R. Conjoint measurement for quantifying judgmental data. Journal of Marketing Research, 1971, 8, 355-363.

Green, P. E. & Wind, Y. New way to measure consumers' judgment. Harvard Business Review, 1975, 53, 107-117.

Hammond, K. R. & Addelman, L. Science: values and human judgment. Science, 1976, 194, 389-396.

Hammond, K. R., Steward, T. R., Brehmer, B., & Steinmann, D. Social judgment theory, in M. Kaplan & S. Schwartz (Eds.). Human Judgment and Decision Process, New York: Academic Press, 1975.

Keeney, R. L. A decision analysis with multiple objectives: the Mexico City Airport. Bell Journal of Economics and Management Science, 1973, 4, 101-117.

Keeney, R. L. A utility function for examining policy affecting salmon in the Skeena River. International Institute for Applied Systems Analysis Research Memo. RM-76-5 (1976), 42 p.

Keeney, R. L. The art of assessing multiattribute utility functions. Organizational Behavior and Human Performance, 1977, 19, 267-310.

Keeney, R. L. Multiplicative utility functions. Operations Research, 1974, 22, 22-34.

Keeney, R. L. & Raiffa, H. Decisions with multiple objectives: preferences and value tradeoffs. New York: Wiley Press, 1971.

Krantz, D., Luce, R. D., Suppes, P., & Tversky, A. Foundations of Measurement (Vol. 1). New York: Academic Press, 1971.

Slovic, P. From Shakespeare to Simon: speculations -- and some evidence -- about man's ability to process information. ORI Res. Mon., 1972, 12(2).

Slovic, P., Fischhoff, B., & Lichtenstein, S. Behavioral decision theory. Annual Review of Psychology, 1977.

Slovic, P. & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.

Tversky, A. On the elicitation of preferences: descriptive and prescriptive considerations, in D. E. Bell, R. L. Keeney, and H. Raiffa (Eds.), Conflicting objectives in decisions. New York: Wiley, 1977.

von Winterfeldt, D. Experimental tests of independence assumptions for risky multiattribute preferences. Research Report 76-8, Social Science Research Institute, University of Southern California, 1976.

Wind, Y & Spitz, L. K. Analytical approach to marketing decisions in health-care organizations. Operations Research, 1976, 24, 972-990.

Yntema, D. B. & Torgerson, W. S. Man-computer cooperation in decisions requiring common sense. IRE Trans. on Human Factors in Elections, 1961, HFE-2, 20-26.

Table 1

Attribute Levels for a 4^5 Orthogonal Design

Consequence	Attributes				
	1	2	3	4	5
1	1*	1	1	1	1
2	1	2	2	3	4
3	1	3	3	4	2
4	1	4	4	2	3
5	2	1	2	2	2
6	2	2	3	4	3
7	2	3	4	3	1
8	2	4	1	1	4
9	3	1	3	3	3
10	3	2	4	1	2
11	3	3	1	2	4
12	3	4	2	4	1
13	4	1	4	4	4
14	4	2	1	2	1
15	4	3	2	1	3
16	4	4	3	3	2

*Level 1 is worst; level 4 is best.

Table 2

Attribute Level Combinations, Stated Holistic Judgments

and Predicted Judgments for Two Orthogonal Designs

Combination	Design 1 1*	2	3	Stated Value	Predicted Value	Design 2 1	2	3	Stated Value	Predicted Value
1	W	W	W	.00	.00	W	W	W	.00	.00
2	W	I	I	.60	.57	I	W	I	.47	.38**
3	W	B	B	.85	.78	B	W	B	.55	.55
4	I	W	I	.47	.38**	I	I	W	.39	.40
5	I	I	B	.75	.73	B	I	I	.65	.70
6	I	B	W	.50	.53	W	I	B	.70	.66
7	B	W	B	.55	.55	B	B	W	.65	.59
8	B	I	W	.60	.46**	W	B	I	.70	.69
9	B	B	I	.82	.89	I	B	B	.85	.84
10	W	B	W	.60		W	B	W	.60	

*Attributes are designated 1, 2, 3. Levels are denoted W, for worst; I, for intermediate; B, for best. Parameter K is estimated using combination 10.

**Stated value differs from predicted value by at least .077, twice the estimated root mean square error.

Table 3

Stated Holistic Judgments, Original Predictions and

Revised Predictions for Two Orthgonal Designs

	Design 1			Design 2		
Combination	Stated Value	Revised Prediction	Original Prediction	Stated Value	Revised Prediction	Original Prediction
1	.00	.00	.00	.00	.00	.00
2	.60	.57	.57	.47	.40	.38
3	.85	.79	.78	.55	.55	.55
4	.47	.40	.38	.39	.38	.40
5	.75	.73	.73	.65	.68	.70
6	.50	.53	.53	.70	.66	.66
7	.55	.55	.55	.65	.57	.59
8	.50*	.43	.46	.70	.70	.69
9	.82	.81	.89	.85	.85	.84

*Revised from .60 by assumption.

CONTRACT DISTRIBUTION LIST
(Unclassified Technical Reports)

2 copies

Director
Advanced Research Projects Agency
Attention: Program Management Office
1400 Wilson Boulevard
Arlington, Virginia 22209

3 copies

Office of Naval Research
Attention: Code 455
800 North Quincy Street
Arlington, Virginia 22217

12 copies

Defense Documentation Center
Attention: DDC-TC
Cameron Station
Alexandria, Virginia 22314

1 copy

DCASMA Baltimore Office
Attention: Mr. K. Gerasim
300 East Joppa Road
Towson, Maryland 21204

6 copies

Director
Naval Research Laboratory
Attention: Code 2627
Washington, D.C. 20375

SUPPLEMENTAL DISTRIBUTION LIST
(Unclassified Technical Reports)

Department of Defense

Director of Net Assessment
Office of the Secretary of Defense
Attention: MAJ Robert G. Gough, USAF
The Pentagon, Room 3A930
Washington, DC 20301

Assistant Director (Net Technical Assessment)
Office of the Deputy Director of Defense
  Research and Engineering (Test and
  Evaluation)
The Pentagon, Room 3C125
Washington, DC 20301

Director, Defense Advanced Research
  Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, ARPA Regional Office (Europe)
Headquarters, U.S. European Command
APO New York 09128

Director, ARPA Regional Office (Pacific)
Staff CINCPAC, Box 13
Camp H. M. Smith, Hawaii 96861

Dr. Don Hirta
Defense Systems Management School
Building 202
Ft. Belvoir, VA 22060

Chairman, Department of Curriculum
  Development
National War College
Ft. McNair, 4th and P Streets, SW
Washington, DC 20319

Defense Intelligence School
Attention: Professor Douglas E. Hunter
Washington, DC 20374

Vice Director for Production
Management Office (Special Actions)
Defense Intelligence Agency
Room 1E863, The Pentagon
Washington, DC 20301

Command and Control Technical Center
Defense Communications Agency
Attention: Mr. John D. Hwang
Washington, DC 20301

Department of the Navy

Office of the Chief of Naval Operations
  (OP-951)
Washington, DC 20450

Office of Naval Research
Assistant Chief for Technology (Code 200)
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research
Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office
Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office
Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office
Attention: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Director, ONR Branch Office
Attention: Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106

Office of Naval Research
Scientific Liaison Group
Attention: Dr. M. Bertin
American Embassy - Room A-407
APO San Francisco 96503

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps (Code RD-1)
Washington, DC 20380

Headquarters, Naval Material Command
 (Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Dean of Research Administration
Naval Postgraduate School
Attention: Patrick C. Parker
Monterey, CA 93940

Superintendent
Naval Postgraduate School
Attention: R. J. Roland, (Code 52R1)
            C^3 Curriculum
Monterey, CA 93940

Naval Personnel Research and Development
  Center (Code 305)
Attention: LCDR O'Bar
San Diego, CA 92152

Navy Personnel Research and Development
  Center
Manned Systems Design (Code 311)
Attention: Dr. Fred Muckler
San Diego, CA 92152

Naval Training Equipment Center
Human Factors Department (Code N215)
Orlando, FL 32813

Naval Training Equipment Center
Training Analysis and Evaluation Group
  (Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Director, Center for Advanced Research
Naval War College
Attention: Professor C. Lewis
Newport, RI 02840

Naval Research Laboratory
Communications Sciences Division (Code 54C
Attention: Dr. John Shore
Washington, DC 20375

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, MD 21402

Chief, Intelligence Division
Marine Corps Development Center
Quantico, VA 22134

Department of the Army

Deputy Under Secretary of the Army
(Operations Research)
The Pentagon, Room 2E621
Washington, DC 20310

Director, Army Library
Army Studies (ASDIRS)
The Pentagon, Room 1A534
Washington, DC 20310

U.S. Army Research Institute
Organizations and Systems Research Laboratory
Attention: Dr. Edgar M. Johnson
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and Systems
  Research Laboratory
U.S. Army Institute for the Behavioral
  and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director, U.S. Army Concepts
  Analysis Agency
8120 Woodmont Avenue
Bethesda, MD 20014

Director, Strategic Studies Institute
U.S. Army Combat Developments Command
Carlisle Barracks, PA 17013

Commandant, Army Logistics Management Center
Attention: DRXMC-LS-SCAD (ORSA)
Ft. Lee, VA 23801

Department of Engineering
United States Military Academy
Attention: COL A. F. Grum
West Point, NY 10996

Marine Corps Representative
U.S. Army War College
Carlisle Barracks, PA 17013

Chief, Studies and Analysis Office
Headquarters, Army Training and Doctrine
  Command
Ft. Monroe, VA 23351

Commander, U.S. Army Research Office
  (Durham)
Box CM, Duke Station
Durham, NC 27706

Department of the Air Force

Assistant for Requirements Development
  and Acquisition Programs
Office of the Deputy Chief of Staff for
  Research and Development
The Pentagon, Room 4C331
Washington, DC 20330

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Commandant, Air University
Maxwell AFB, AL 36112

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Deputy Chief of Staff, Plans, and
  Operations
Directorate of Concepts (AR/XOCCC)
Attention: Major R. Linhard
The Pentagon, Room 4D 1047
Washington, DC 20330

Director, Advanced Systems Division
  (AFHRL/AS)
Attention: Dr. Gordon Eckstrand
Wright-Patterson AFB, OH 45433

Commander, Rome Air Development Center
Attention: Mr. John Atkinson
Griffis AFB
Rome, NY 13440

IRD, Rome Air Development Center
Attention: Mr. Frederic A. Dion
Griffis AFB
Rome, NY 13440

HQS Tactical Air Command
Attention: LTCOL David Dianich
Langley AFB, VA 23665

Other Government Agencies

Chief, Strategic Evaluation Center
Central Intelligence Agency
Headquarters, Room 2G24
Washington, DC 20505

Director, Center for the Study of
  Intelligence
Central Intelligence Agency
Attention: Mr. Dean Moor
Washington, DC 20505

Mr. Richard Heuer
Methods & Forecasting Division
Office of Regional and Political Analysis
Central Intelligence Agency
Washington, DC 20505

Office of Life Sciences
Headquarters, National Aeronautics and
  Space Administration
Attention: Dr. Stanley Deutsch
600 Independence Avenue
Washington, DC 20546

Other Institutions

Department of Psychology
The Johns Hopkins University
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Stanford University
Attention: Dr. R. A. Howard
Stanford, CA 94305

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A. D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Brunel University
Attention: Dr. Lawrence D. Phillips
Uxbridge, Middlesex UB8 3PH
England

Decision Analysis Group
Stanford Research Institute
Attention: Dr. Miley W. Merkhofer
Menlo Park, CA 94025

Decision Research
1201 Oak Street
Eugene, OR 97401

Department of Psychology
University of Washington
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Department of Electrical and Computer
  Engineering
University of Michigan
Attention: Professor Kan Chen
Ann Arbor, MI 94135

Department of Government and Politics
University of Maryland
Attention: Dr. Davis B. Bobrow
College Park, MD 20747

Department of Psychology
Hebrew University
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. Andrew P. Sage
School of Engineering and Applied
  Science
University of Virginia
Charlottesville, VA 22901

Professor Raymond Tanter
Political Science Department
The University of Michigan
Ann Arbor, MI 48109

Professor Howard Raiffa
Morgan 302
Harvard Business School
Harvard University
Cambridge, MA 02163

Department of Psychology
University of Oklahoma
Attention: Dr. Charles Gettys
455 West Lindsey
Dale Hall Tower
Norman, OK 73069

Institute of Behavioral Science #3
University of Colorado
Attention: Dr. Kenneth Hammond
Room 201
Boulder, Colorado 80309

Decisions and Designs, Incorporated
Suite 600, 8400 Westpark Drive
P.O. Box 907
McLean, Virginia 22101

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 001922-2-T	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER Research rept.
4. TITLE (and Subtitle) Validation and Error in Multiplicative Utility Functions.		5. TYPE OF REPORT & PERIOD COVERED Oct 77 - Dec 78 Technical -10/77--12/78
		6. PERFORMING ORG. REPORT NUMBER SSRI 78-2
7. AUTHOR(s) F. Hutton Barron		8. CONTRACT OR GRANT NUMBER(s) Prime Contract N00014-76-C-0074 Subcontract 76-030-0715
9. PERFORMING ORGANIZATION NAME AND ADDRESS Social Science Research Institute University of Southern California Los Angeles, California 90007		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd Arlington, Virginia 22209		12. REPORT DATE December 78 13. NUMBER OF PAGES 50
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Decisions & Designs, Inc. 8400 Westpark Drive, Suite 600 McLean, Virginia 22101 (under contract from Office of Naval Res.		15. SECURITY CLASS. (of this report) unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

12 51 p.

17. DISTRIBUTION STATEMENT (of the abstract entered in block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

holistic assessment methods   multiplicative function
utility function   additive function
scaling constants
orthogonal

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this report an approach to the concept of error in utility assessment is proposed. Three components of error are considered and each component is related to four separate elicitation methods—all in the context of a general multiplicative multiattribute utility model. The methods are Keeney-Raiffa (1976) procedure, SMART (Edwards, 1977), a social judgment theory (SJT) based regression model (Hammond, Stewart, Brehmer and Steinmann, 1975) and

DD FORM 1473   EDITION OF 1 NOV 65 IS OBSOLETE
1 JAN 73   S/N 0102-014-6601

390 664

a new method called Holistic Orthogonal Parameter Estimation or HOPE (Barron and Person, 1978).

If a general multiplicative model can be assumed to be an appropriate representation of the decision maker's basic preference structure, error can occur in the direct estimation of the scaling constants and univariate utility functions for decomposition methods (Keeney-Raiffa and SMART), or in the holistic assessments for holistic methods (SJT and HOPE). Individual estimates may be merely noisy or may be fundamentally incorrect. Furthermore, the utility model may be incorrectly specified; for example, an additive model, rather than a multiplicative model, may be used. The four assessment methods are considered in conjunction with errors of each kind.

The most serious error-method combination is the case of a substantial degree of error occurring in a single holistic judgment which is being used in a HOPE procedure. This concern leads to a major emphasis of this report--and expanded HOPE procedure used in conjunction with a convergent validation strategy to estimate error in individual holistic judgments and thus guide consistency checks.

The discussion is organized into four sections. The HOPE procedure is summarized in Section I. In Section II, three components of assessment error are considered in conjunction with the four elicitation procedures. In Section III, an expanded HOPE procedure for detecting judgment error and guiding consistency checks is proposed. In Section IV, application considerations are outlined.